

End User License Agreement - NICE Dataset

This dataset consists of image-comment pairs. We filtered both the images and comments to remove sensitive content such as adult or pornographic content, racy and gory content, non-English language, ethnic-religious content, and some sensitive content (including people's name, documents invoices, bills, financial reports) or other potentially offensive or contentious material (including inappropriate references to violence, crime and illegal substances). We keep no more than 5 corresponding comments in each thread, and no more than 6 different dialogue threads for the same image. We also filter out image-comments pairs that do not have affect features or dialogue topics from dialogue thread. The final set includes 6,550,526 training samples, 100,000 validation samples, and 70,000 testing samples.

Please add a short description of what you will use the data for:

Terms of Use:

- Use of this dataset is governed by the CC0 1.0 Universal license.

If you use the dataset please cite:

NICE: Neural Image Commenting with Empathy, K. Chen et al.

Name

Email Address

Signature

Institution

Title

Date
